–Supplemental Document–
On Transferability of Histological Tissue Labels in Computational Pathology

Mahdi S. Hosseini[1 [0000−0002−9147−0731]], Lyndon Chan[1 [0000−0002−1185−7961]], Weimin Huang[1 [0000−0001−6041−9848]], Yichen Wang[1 [0000−0001−5969−3233]], Danial Hasan[1 [0000−0003−0620−4402]], Corwyn Rowsell[2,3 [0000−0002−0604−638X]], Savvas Damaskinos[4 [0000−0001−6030−1823]], and Konstantinos N. Plataniotis[1 [0000−0003−3647−5473]]

[1] Department of Electrical & Computer Engineering, University of Toronto
[2] Division of Pathology, St. Michaels Hospital, Toronto, ON, M4N 1X3, Canada
[3] Department of Laboratory Medicine and Pathobiology, University of Toronto
[4] Huron Digital Pathology, St. Jacobs, ON, N0B 2N0, Canada
{mahdi.hosseini, lyndon.chan, cheryl.huang, yichenk.wang, danial.hasan, kostas.plataniotis}@utoronto.ca
https://github.com/mahdihosseini/HistoLabelTransfer/

1 ADP Label Set Modifications

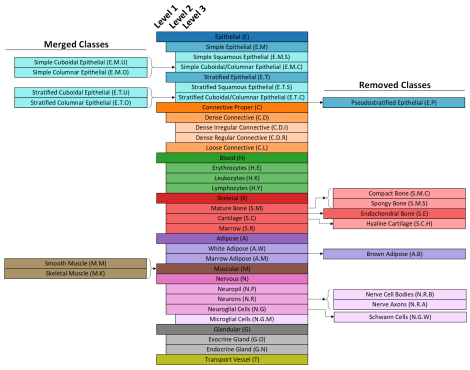


Fig. 1. The *ADP-flat* database is a modification of the original *ADP* database, but avoids some notable mislabeling cases by either merging (left) or omitting (right) frequently-mislabeled classes.

The Atlas of Digital Pathology (ADP) [3] is a database of 17668 digital pathology patch images extracted from 100 healthy slides from the same medical institution scanned with a TissueScope LE1.2 at 0.25μm/px resolution. Each 1088×1088 patch in the original release was annotated with up to 42 hierarchical tissue types.

The database was shown to be well-annotated by training state-of-the-art CNNs on the labels, although several classes were frequently mislabeled. To mitigate the negative effects this might have on training, we modified the original label set by (1) merging classes noted to be mislabeled as each other and (2) removing rarely-occurring classes. We selected the ADP database as the source domain for domain adaptation because, unlike other available computational pathology databases, it contains data from diverse organs, utilizes a broadly-applicable label set, and overwhelmingly contains healthy tissue. Intuitively, this should make it possible to apply to datasets containing observed organs, observed labels, and possibly even detect tissue abnormalities.

2 Histological Label Transferring Implementation

2.1 Dataset Details

Below, we provide extended details for each of the datasets evaluated in the main paper, as well as the BACH dataset which was not included for reasons of space.

CRC–The ColoRectal Cancer dataset [4] consists of 100000 patch images (NCT-CRC-HE-100K-NONORM, or CRC-100K), each labeled with one of nine tissue type labels. Five labels are largely healthy tissue types (e.g. 'ADI' for adipose tissue, 'LYM' for lymphocytes), two are disease types (i.e. 'STR' for cancerous stroma, 'TUM' for cancerous epithelia), and two are non-tissue types (i.e. 'BACK' for background, 'DEB' for debris). The images are sized 224×224 and scanned at 0.5μm/px, so they were resized by a factor of $\frac{0.5}{1} = 0.5$ to 112×112 and symmetric padded by 80 pixels on each side. An additional set of 7000 patch images color normalized using Macenko *et al.*'s method [6] was also provided (CRC-VAL-HE-7K, or CRC-7K) but as these are color-normalized, we only use the 100000 un-normalized images for evaluation - the 7000 normalized set results are provided below.

HMT–The Histology Multiclass Texture dataset [5] consists of 5000 patch images, each labeled with one of eight tissue type labels. Five of these labels are largely healthy tissue types (e.g. '02_STROMA' for stroma, '04_LYMPHO' for lymphocytes), while one is a disease type (i.e. '01_TUMOR'), and two are non-tissue types (i.e. '05_DEBRIS', '08_EMPTY'). The images are sized 150×150 and scanned at 0.495μm/px, so they were resized by a factor of $\frac{0.495}{1} = 0.495$ to 74×74 and symmetric padded by 99 pixels on each side. Ten additional slide images (sized 5000×5000) are also provided for validation, but we use the 5000-image training set for evaluation.

GlaS–The Gland Segmentation (GlaS) challenge dataset [7] consists of 85 training images and 80 test images. Each image is labeled with one of five colorectal cancer grades - in order of worsening condition, they are: 'healthy', 'adenomatous', 'moderately differentiated', 'moderately-to-poorly differentiated', and 'poorly differentiated'. All images are scanned at 0.62005μm/px (Zeiss MIRAX MIDI) and the median image is sized 775×522, so they were resized by a factor

of $\frac{0.62005}{1} = 0.62005$ to 481×324 and a minimally-overlapping set of 2×2 crops was extracted (for the median case). We combined both the training and test sets for evaluation.

PatchCamelyon–The PatchCamelyon dataset [8] consists of 32768 patch images extracted from the WSI scans of the original Camelyon16 challenge dataset [2] and labeled as tumorous if the central 32×32 region contains at least one pixel of tumor; it is labeled normal otherwise. The images are sized 96×96 and scanned at 0.243μm/px (Pannoramic 250 Flash II; 3DHISTECH), so they were resized by a factor of $\frac{0.243}{1} = 0.243$ to 23×23 and symmetric padded by 125 pixels on each side. We evaluate on the test annotations of this dataset.

BACH–The Grand Challenge on Breast Cancer Histology Images dataset [1] consists of 400 training images and 100 test image. Each image is labeled with one of four breast cancer classes - in order of worsening condition, they are: 'Normal', 'Benign', 'InSitu', and 'Invasive'. The images are sized 2048×1536 and scanned at 0.42μm/px, so they were resized by a factor of $\frac{0.42}{1} = 0.42$ to 860×645 and a minimally-overlapping set of 4×3 crops was extracted. We used only the training set for evaluation.

3 Domain Adaptation Experiments, Extended Results

3.1 CRC

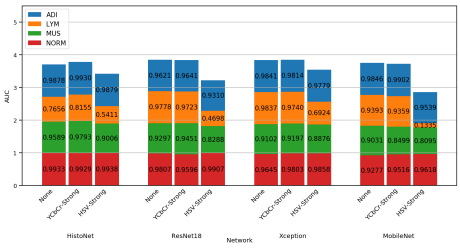


Fig. 2. CRC-100K (non-normalized): per-class AUC for four network architectures trained with three color augmentation methods each.

3.2 HMT

3.3 GlaS

3.4 PatchCamelyon

3.5 BACH

4 Visualizing Some Failure Cases

See Figure 10 for a visualization of some true positive and false positive predictions in the CRC and GlaS sets.

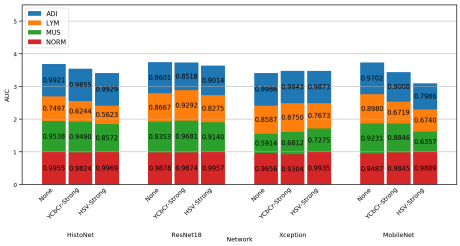


Fig. 3. CRC-7K (color normalized): per-class AUC for four network architectures trained with three color augmentation methods each.

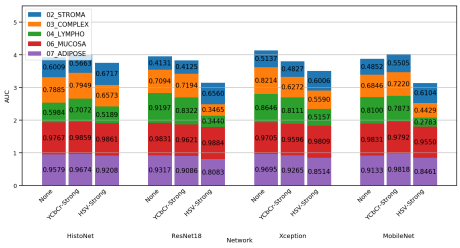


Fig. 4. HMT: per-class AUC for four network architectures trained with three color augmentation methods each.

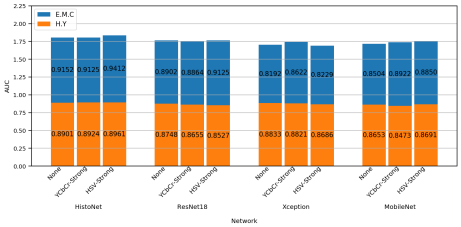


Fig. 5. GlaS: disease detection AUC for four network architectures trained with three color augmentation methods each, using the E.M.C and H.Y classes.

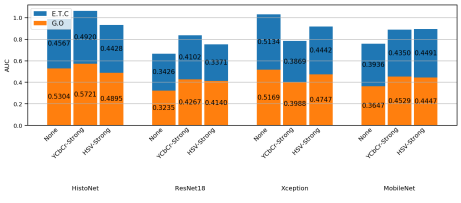


Fig. 6. PatchCamelyon: disease detection AUC for four network architectures trained with three color augmentation methods each, using the E.T.C and G.O classes.

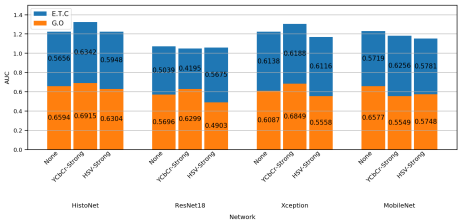


Fig. 7. BACH: disease detection AUC for four network architectures trained with three color augmentation methods each, using the E.T.C and G.O classes.

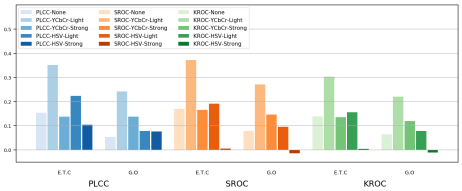


Fig. 8. Correlation between select inverse HistoNet confidence scores (E.T.C and G.O) and binarized disease classes in BACH (normal/tumor): PLCC (Pearson Linear Correlation Coefficient), SROC (Spearman Rank Order Coefficient), and KROC (Kendall Rank Order Coefficient).

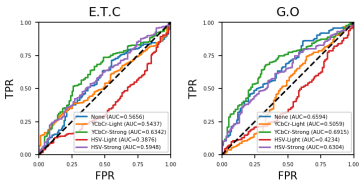


Fig. 9. ROC curves of HistoNet in PatchCamelyon trained with different color augmentation methods and evaluated on binary disease detection using two tissue types: E.M.C and H.Y.

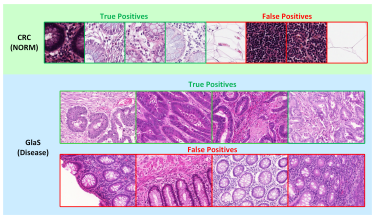


Fig. 10. True and false positive predictions in tissue classification (NORM class in CRC) and disease detection (in GlaS).

4.1 PatchCamelyon

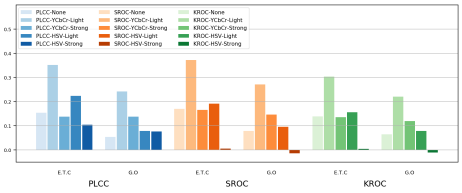


Fig. 11. Correlation between select inverse HistoNet confidence scores (E.T.C and G.O) and binarized disease classes in PatchCamelyon (normal/tumor): PLCC (Pearson Linear Correlation Coefficient), SROC (Spearman Rank Order Coefficient), and KROC (Kendall Rank Order Coefficient).

Statistical Correlation with Disease Classes. For PatchCamelyon, two ADP classes - stratified cuboidal/columnar epithelium (E.T.C) and exocrine gland (G.O) - had the highest correlation with the presence of tumors. Figure 12 shows that this correlation is most significant when the network is trained with YCbCr-Light augmentation. Although the correlation values are relatively low, the consistency of correlation is still remarkable because the patches are extracted from breast tissue, which is not found in ADP. The selection of stratified cuboidal/columnar epithelium and exocrine gland is intriguing because the metastatic tumor cells are epithelial in nature but result in low scores for E.T.C and G.O.

Binary Disease Detection. When the E.T.C and G.O scores are used, HistoNet fails to properly detect disease in PatchCamelyon unless YCbCr-Light is applied, as shown in Figure 12 ([8] reports an AUC of 0.963). This justifies the utility of color augmentation, since reasonable prediction are not possible otherwise and the usefulness of our approach in detecting disease in organs not seen in the source domain.

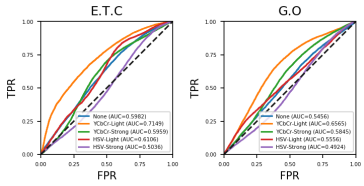


Fig. 12. ROC curves of HistoNet in PatchCamelyon trained with different color augmentation methods and evaluated on binary disease detection using two tissue types: E.M.C and H.Y.

5 Disease Analysis on WSI Level

In this section, we provide visual results on WSI analysis. For the experiment, two Colon tissue organs are selected as shown in Figure 13. Both WSIs are mosaiced into multiple image patches and fed to the HistoNet model with $1\mu m$/pixel resolution for HTT prediction. Accordingly, image patches are down sized to 272×272 and normalized before feeding to HistoNet. We construct the heatmap for four different color augmentation models shown in Figure 14 corresponding to inverse prediction score of Stratified Cuboidal/Columnar Epithelial (E.T.C) tissue type. In the following subsections, the pathologist's evaluation on each organ are provided.

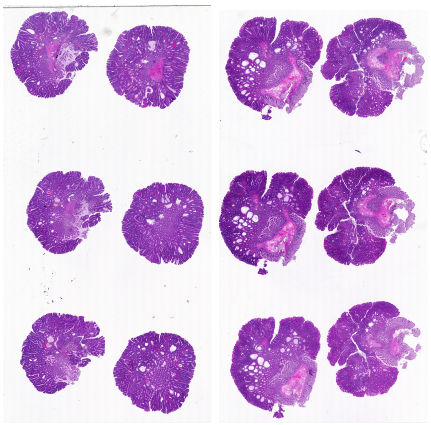
5.1 Pathologist's Diagnosis on Colon-1

This WSI depicts an adenomatous polyp of the colon, shown in Figure 13(a). The majority of the epithelium in this slide is abnormal (neoplastic, precancerous), but there is an area of muscularis mucosa and normal epithelium at area where the polyp was removed. The heatmaps in Figure 14 (first row) show a high probability of abnormality in the areas of adenomatous epithelium (yellow, orange, and red), and indicate a low probability of abnormality in the regions with muscularis mucosa and normal epithelium (blue). The HSV Strong protocol appears to show the strongest correlation with histologic findings, followed by HSV Light, then YCbCr Light. The YCbCr Strong protocol shows the least correlation (while it still correctly indicates the normal areas, it appears to be less sensitive in identifying areas of abnormality compared to the other methodologies).

5.2 Pathologist's Diagnosis on Colon-2

This WSI is also from a section of an adenomatous polyp, shown in Figure 13(b). Similar to the previous case, the majority of the columnar epithelium on the slide shows histologic abnormality, but there are regions of normal tissue (muscularis mucosa) and normal columnar epithelium. In the associated heatmaps shown in Figure 14 (second row), the areas of normal muscularis mucosa are marked as

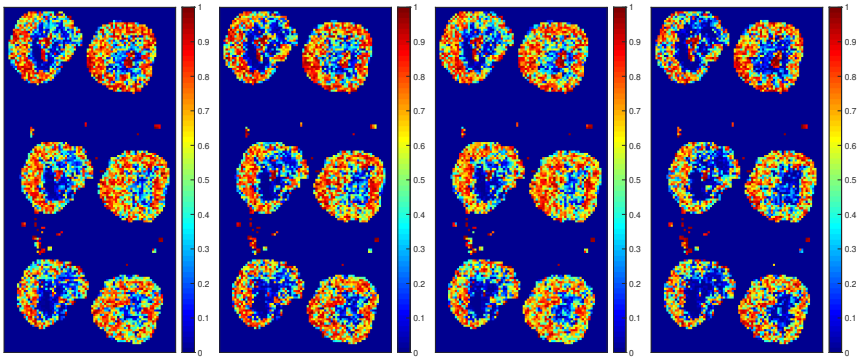
'abnormal' with a predominance of red. Areas of normal mucosa are appropriately marked blue, while the abnormal adenomatous mucosa shows a variable pattern. It appears that for this slide, the AI tool is marking non-epithelial tissues as abnormal. Whereas, this is expected as we demonstrate the abnormality heatmap solely based on Epithelial section of the HTT. The protocol with the best correlation with epithelial abnormality in this series is YCbCr light, followed by HSV strong, HSV light. YCbCr strong shows the least correlation.



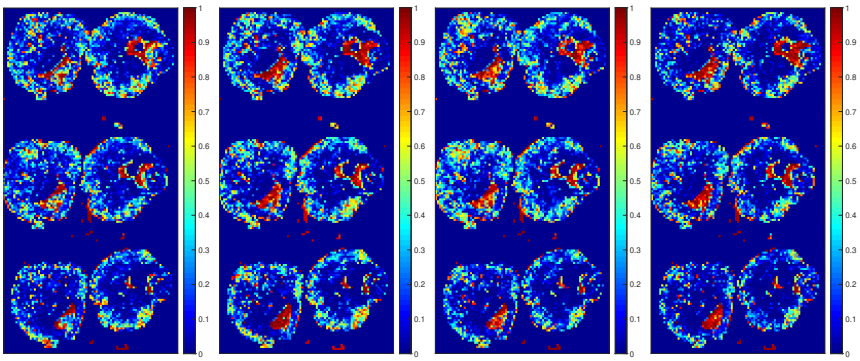
(a) Colon-1: Adenomatous Polyp (b) Colon-2: Adenomatous Polyp

Fig. 13. Two Colon tissue organs are selected for processing and diagnosed by pathologist.

There is overall much better correlation on the first case i.e. Color-1. It is interesting in the first case i.e. Color-1 that muscularis mucosa correlated with normal on the heatmap, but did the opposite in the second case i.e. Color-2.



(a) Colon-1: HSV-Light (b) Colon-1: HSV-Strong (c) Colon-1: YCbCr-Light (d) Colon-1: YCbCr-Strong



(e) Colon-2: HSV-Light (f) Colon-2: HSV-Strong (g) Colon-2: YCbCr-Light (h) Colon-2: YCbCr-Strong

Fig. 14. Heatmaps of Abnormality based on the inverse prediction of Stratified Cuboidal/Columnar Epithelial (E.T.C). The first and the second rows correspond to the Colon-1 and Colon-2 organs, accordingly. For visual overlay, please refer to Figure 13 for the original WSI scans.

References

1. Aresta, G., Araújo, T., Kwok, S., Chennamsetty, S.S., Safwan, M., Alex, V., Marami, B., Prastawa, M., Chan, M., Donovan, M., et al.: Bach: Grand challenge on breast cancer histology images. Medical image analysis (2019)
2. Bejnordi, B.E., Veta, M., Van Diest, P.J., Van Ginneken, B., Karssemeijer, N., Litjens, G., Van Der Laak, J.A., Hermsen, M., Manson, Q.F., Balkenhol, M., et al.: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. Jama **318**(22), 2199–2210 (2017)
3. Hosseini, M.S., Chan, L., Tse, G., Tang, M., Deng, J., Norouzi, S., Rowsell, C., Plataniotis, K.N., Damaskinos, S.: Atlas of digital pathology: A generalized hierarchical histological tissue type-annotated database for deep learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 11747–11756 (2019)
4. Kather, J.N., Krisam, J., Charoentong, P., Luedde, T., Herpel, E., Weis, C.A., Gaiser, T., Marx, A., Valous, N.A., Ferber, D., et al.: Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. PLoS medicine **16**(1), e1002730 (2019)
5. Kather, J.N., Weis, C.A., Bianconi, F., Melchers, S.M., Schad, L.R., Gaiser, T., Marx, A., Zöllner, F.G.: Multi-class texture analysis in colorectal cancer histology. Scientific reports **6**, 27988 (2016)
6. Macenko, M., Niethammer, M., Marron, J.S., Borland, D., Woosley, J.T., Guan, X., Schmitt, C., Thomas, N.E.: A method for normalizing histology slides for quantitative analysis. In: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. pp. 1107–1110. IEEE (2009)
7. Sirinukunwattana, K., Pluim, J.P., Chen, H., Qi, X., Heng, P.A., Guo, Y.B., Wang, L.Y., Matuszewski, B.J., Bruni, E., Sanchez, U., et al.: Gland segmentation in colon histology images: The glas challenge contest. Medical image analysis **35**, 489–502 (2017)
8. Veeling, B.S., Linmans, J., Winkens, J., Cohen, T., Welling, M.: Rotation equivariant cnns for digital pathology. In: International Conference on Medical image computing and computer-assisted intervention. pp. 210–218. Springer (2018)